DOCUMENT RESUME

ED 222 496 TM 820 068

AUTHOR	Burstein, Leigh; Miller, Michael David
TITLE	Report of Project Activities [and] Selected Topics in Regression-Based Analyses of Multilevel Educational Data. Measurement and Methodology, Work Unit 3: Analysis of Multilevel Data.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Jun 78
NOTE	49p.; Pages 37-44 are of marginable legibility due to broken type.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Data Analysis; *Evaluation Methods; *Group Testing; Measurement Techniques; Quasiexperimental Design; *Regression (Statistics); Research Methodology; *Research Projects; Statistical Analysis
IDENTIFIERS	Analytical Methods; *Center for the Study of Evaluation CA; *Unit of Analysis Problems

ABSTRACT

The activities of the Multilevel Analysis Project of the Center for the Study of Evaluation from February 1 to June 1, 1978 are summarized. Twelve project staff activities are listed, including papers on alternative analytical models for identifying educational effects, conference presentations, and grant proposals. A report on empirical comparisons for a large data set is included: "Selected Topics in Regression-Based Analyses of Multilevel Educational Data." Within the domain of regression-based analyses of multilevel data from quasi-experiments and field studies in educational research and evaluation, the question of choice of unit of analysis or analytical model is discussed. The use of within-group slopes as indices in between-group analyses and the estimation of within-group dependency and its role in analyses of multilevel data are considered with illustrative empirical data. A trend to meta-analyses developing in educational evaluation is shown, and it is recommended that alternative analysis approaches be applied to a wide variety and sizeable number of actual data sets, each with potentially differing sets of inadequacies. Both the methods and the influence of data limitations on methods need further study. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222496

DELIVERABLE: June 1978

MEASUREMENT AND METHODOLOGY

WORK UNIT 3

Analysis of Multilevel Data

Leigh Burstein
Project Director

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

REPORT OF PROJECT ACTIVITIES
Leigh Burstein

SELECTED TOPICS IN REGRESSION-BASED ANALYSES OF MULTILEVEL EDUCATIONAL DATA
Leigh Burstein, Michael David Miller

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. S. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 068

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

June 1978

"The activity which is the subject of this report was supported in part or in whole by the National Institute of Education, Department of HEW. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred."

Final Report

for period ending June 1, 1978

Multilevel Analysis Project

Leigh Burstein, Project Director

This report describes the activities of the Multilevel Analysis Project for the period February 1 to June 1, 1978; includes material submitted to fulfill the obligation for Product (3); and summarizes work since the beginning of the project.

Activities for February 1 to June 1, 1978

Staff of the Multilevel Analysis Project have done the following project related work during the months of February through May.

(1) Dr. Burstein, with the assistance of Mr. Miller, wrote a paper entitled "Alternative Analytical Models for Identifying Educational Effects: Where are We?" Dr. Burstein presented this paper as part of a project-organized symposium at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 27, 1978.

(2) Dr. J. Ward Keesling (Systems Development Corporation) and Dr. David Rogosa (University of Chicago), project consultants, also presented papers as part of the Symposium on Alternative Analytical Models for Identifying Educational Effects. Other symposium participants were Dr. Nancy Karweit (CSOS, John Hopkins) who presented a paper co-authored by Dr. James Fennessey and Ms. Denise Daiger, Mr. Carlyle Maw (NIE) who served as chairman, and Dr. Robert Linn (University of Illinois) and Dr. Gene V. Glass, who served as discussant. Revised versions of Dr. Keesling's and Dr. Rogosa's presentations will eventually be submitted as products of the Multilevel Analysis Project and will appear in the

CSE Report Series.

(3) Mr. Frank Capell, doctoral candidate at Stanford University and co-project director of CSE's CSA evaluation and will be working with the project director on all aspects of its conceptual work and report preparation.

(4) Dr. Burstein's presentation at the CSE Winter Measurement and Methodology Conference was revised for inclusion in the Conference Proceedings. A 500-word abstract of the paper was drafted for inclusion in a future issue of Evaluation Comment on the conference.

(5) Dr. Burstein, with the assistance of Mr. Capell, who joined the project staff part-time, wrote a proposal for the Multilevel Analysis Project for an NIE 18 month grant. He also drafted the description of CSE's Quantitative Analysis Group which includes projects involving Dr. Rand Wilcox and Dr. Jenny Yeh.

(6) Dr. Burstein drafted a long abstract describing Project activities for inclusion in the current Evaluation Comment on work being conducted at CSE.

(7) Dr. Burstein and Mr. Miller revised the paper presented earlier at the Institute for Research on Teaching, Michigan State University. This paper is being submitted to fulfill the committment for Product ③ (Report on empirical comparisons for large data set).

(8) Dr. Burstein, Mr. Miller, and Dr. Fischer (project director of the Bilingual Prediction Study) continued work on a paper entitled "Social Policy and School Effects: A Cross-National Comparison." Dr. Burstein will present this paper at the World Congress of Sociology Meetings in Uppsala, Sweden. A revised version will be submitted as a product of the 18 month grant and will be submitted to the journal Sociology of Education for consideration for publication.

(9) Dr. Burstein revised "Alternative Approaches for Assessing Differences Between Group and Individual-Level Regression Coefficients" (Product ①). This paper will appear in the August 1978 issue of Sociological Methods and Research.

(10) Dr. Burstein and Dr. Robert L. Linn have been revising "Alternative Analytical models for Identifying Educational Effects in the Presence of Heterogeneous Within Group Relations of Outcome to Input" (Product 2) with the assistance of Dr. Rogosa and Mr. Capell. The revised version will be resubmitted to the Journal of Educational Statistics at the editor's request.

(11) Mr. Miller has been completing file clean-up for other countries and other populations of the IEA Six Subject Survey. The project now has access to data from all countries and all populations.

(12) Mr. Miller has also continued to generate total, between, and within regression analyses for IEA countries and "means, s.d. and slopes as group outcomes" for the same populations and countries. These empirical data will serve as part of the data to be analyzed and reported on under the next 18-month plan.

Documents Submitted in Fulfillment of Product 3

Documents submitted in filfillment of Product ③ is attached as Appendix A of this report. Product ③ was originally due July, 1978 and is supposed to include empirical comparisons of selected analysis approaches on a large data set. The attached document is:

(1) Burstein, L. and Miller, M.D. Selected Topics in Regression-Based Analyses of Multilevel Educational Data

-- This is a revised version of a paper originally presented at the

Institute for Research on Teaching, Michigan State University. The paper considers (a) the general question of the choice of an appropriate analytical model for multilevel data, (b) the use of within-group slopes as indices in between-group analyses, and (c) the estimation of within-group dependency and its role in analyses of multilevel data. Empirical comparisons of total, between and within-school regression analyses for the U.S. and IEA data on Sweden are reported and discussed. Also empirical data on within-class and within-school slopes from the ETS BTES Study (McDonald and Elias, 1976), a study by Marco (JEM, 1974) and the IEA Science data for the U.S. are presented. The case of within-school slopes as outcomes for the IEA data is contrasted empirically with the use of school means and school standard deviations. The types of analyses presented represent traditional single-level analyses (between-student (total), between-school, and pooled within-school) and multilevel approaches of the kind advocated by Cronbach (1976; Cronbach and Webb, 1975) and by Burstein and Linn (1976; 1977).

Summary of All Activities of Multilevel Analysis Project

The Multilevel Analysis project became a CSE project in February, 1977. Since its origination, project staff have written the papers and made the presentations listed below.

PAPERS

(1) Burstein, L. Alternative Approaches for Assessing Differences Between Group and Individual-Level Regression Coefficients, to appear in Sociological Methods and Research, August 1978 (Product ①, see 1977 Final Report).

(2) Burstein, L. and Linn, R.L. Alternative Analytical Models for Identifying Teacher Effects in the Presence of Heterogeneous Within-Group Relations of Inputs to Outcomes -- (Revised version of "The Identification of Teacher Effects in Presence of Heterogeneous Within-Class Relations of Inputs to Outcomes"); submitted as Product ②; a revised version to be resubmitted to Journal of Educational Statistics.

(3) Linn, R.L. and Burstein, L. Descriptors of Aggregates, Product ④ (Note this product was not due until November 1978; Further work is being done in this area under the new 18-month plan). A revised version is to be submitted to either Review of Educational Research, Journal of Educational Measurement, or Educational and Psychological Measurement.

(4) Burstein, L. Secondary Analysis: An Important Resource in Educational Research and Evaluation, published in Educational Researcher, May 1978.

(5) Burstein, L. Analyzing Multilevel Educational Data: The Choice of an Analytical Model Rather than a Unit of Analysis -- to appear in Proceedings of CSE Winter Measurement and Methodology Conference.

(6) Burstein, L. and Miller, M.D. Selected Topics in Regression-Based Analysis of Multilevel Data in Large-Scale Educational Research -- Submitted as Product ③.

(9) Symposium on Alternative Analytical Models for Identifying Educational Effects. Presenters: L. Burstein, N. Karweit, J.W. Keesling, D. Rogosa; Chair: C.W. Maw; Discussants: G.V. Glass and R.L. Linn (M.D. Miller co-authored L. Burstein's paper).

Selected Topics in Regression-Based Analyses of Multilevel Educational Data

Leigh Burstein and Michael David Miller[1]

Department of Education

and

Center for the Study of Evaluation

University of California, Los Angeles

[1] An earlier version of this paper was presented at the Institute for Research on Teaching, Michigan State University, East Lansing, Michigan, December 10, 1977.

This paper focuses on three topics within the domain of regression-based analyses of multilevel data from quasi-experiments and field studies in educational research and evaluation. The paper begins with a discussion of the general question of choice of unit of analysis or what may be the more appropriate question of choice of analytical model. After discussing this issue and providing empirical illustrations of the importance of knowing the question of interest, two additional topics will be considered: the use of within-group slopes as indices in between-group analyses and the estimation of within-group dependency and its role in analyses of multilevel data. These latter topics reflect important substantive concerns in school-based non-experimental investigations.

Overall, we believe that the major technical complication in the analysis of multilevel data from quasi-experiments and field studies is the inability of educational researchers to develop adequate theories about educational processes within groups (classrooms and schools) and to develop adequate methodology for analyzing the educational effects of such processes. The material presented here reflects an attempt to systematize the investigation of two important indices of within-group processes.

Choice of Units of Analysis and/or Choice of Analytical Model

Efforts to identify the effects of education (e.g., Coleman, et al., 1966) on pupil performance have suffered from the complications caused by the multilevel character of educational data. Schools are aggregates of their teachers, classrooms and pupils, and classrooms are aggregates of the persons and processes within them. This being the

case, the effects of education can exist both between and within the units at each level of the educational system. Yet the majority of studies of educational effects have restricted attention to either overall between-student, between-class, or between-school analyses.

Cronbach (1976) argued that the majority of studies of educational effects carried out thus far conceal more than they reveal, and that "the established methods have generated false conclusions in many studies" (p. 1). His concern is foreshadowed in the educational literature by the exchange among Wiley, Bloom, and Glaser as recorded in Wittrock and Wiley (1970), and by Haney's (1974) review of the units of analysis problems encountered in the evaluation of Project Follow Through.

Research on the differences between multiple regression models at different levels of aggregation (Blalock, 1964; Burstein, 1974, 1975a, 1975b, 1976a, 1977b; Burstein and Knapp, 1975; Hannan, 1971a, 1971b; Hannan and Burstein, 1974; Hannan and Young, 1976a; Hannan, Young, and Nielsen, 1975; Feige and Watts, 1972) and on the analyses of school effects at different levels (Burstein, 1975a, 1977a, 1977c; Burstein and Knapp, 1975; Burstein, Fischer, and Miller, 1977; Burstein and Smith, 1977; Comber and Keeves, 1973; Hannan, Freeman, and Meyer, 1976; Keesling and Wiley, 1974; Peaker, 1975) indicates that (a) there are substantial differences in the magnitudes of regression coefficients across levels for specific models; (b) different variables enter the models at different levels; and (c) aggregation generally inflates the estimated effects of pupil background and decreases the likelihood of identifying teacher and classroom characteristics that are effective. The results cited above are not very comforting for the researcher who wishes to draw conclusions about educational processes at one level but is constrained to analysis at a different level.

When faced with the analysis of multilevel data, most researchers have tried to make a choice among alternative units of analysis on the basis of theory or statistical considerations. Unfortunately, those who resort to theory either reject plausible alternative models (Brophy, 1975; Bloom in Wittrock and Wiley, 1970; Stebbins et al., 1977; Wiley in Wittrock and Wiley, 1970) or find themselves unable to choose (Cline et al., 1974; Haney, 1974; McBee and Fortune, 1977; Walberg and Singh, 1974). Picking the appropriate unit on the basis of statistical considerations can also leave the choice unresolved due to competing alternatives (Burstein, 1975a; Burstein and Smith, 1977; Glendening, 1976a; Haney, 1974). Haney (1974) has elaborated the range of alternative considerations in the context of the evaluation of Project Follow Through. He cites four general types: the purpose of the evaluation (questions to be addressed), the evaluation design (nature of treatments, independence of units and treatment effects, appropriate size), statistical considerations (reliability of measures, degrees of freedom, analysis techniques), and practical considerations (missing data, policy research, multiple year comparisons, economy). Haney was unable to choose among units because the purpose of the evaluation dictated the child as the unit but the unit of treatment was the classroom; moreover, the multiyear character of Follow Through made classrooms impractical as units of analysis. And, since there was no random assignment at any level and the comparison children were not equivalent to treatment children, these considerations offered no relief.

Apparently, thinking of multilevel analyses simply as problems in the choice of a unit of analysis is inadequate. Phenomena of importance occur at all levels and need to be described and subjected

to inference-making (Burstein and Linn, 1976, 1977; Cronbach, 1976). Once again, Haney's arguments are succinct and to the point:

> Investigators ought to have a strong bias for studying various properties of the educational system at the level at which they occur; . . . variation in attributes of interest ought to be studied at those levels (or between those units) at which it does (or is expected to) occur. . . . If the hypotheses are explicitly stated in terms of mathematical models, the impact of shifting levels of analysis from one unit of analysis to another will be much more easily assessed than if they are not (1974, pp. 96-97).

These arguments cited by Haney serve as justification for the research we describe throughout this paper.

Decomposition into Between-Group and Within-Group Effects

A variety of competing points can be cited as traditional justification for the choice of either pupils or groups (classrooms, schools, etc.) as the appropriate unit of analysis in studies of educational effects. Generally arguments cited are compelling and virtually irreconcilable if a choice of either pupil or group as the only unit is required. The multilevel character of educational data warrants analytical strategies tailored to the identification of educational effects at and within each level of the educational system. Moreover, the complexity of the choice depends on the type of study being conducted as well as the types of outcomes and processes under investigation.

Even in the simplest case, once the existence of specific group membership is acknowledged (e.g., instruction from a specific teacher), any measure that varies over pupils can be decomposed into its between-

group and within-group components. For example, if we consider the posttest or outcome performance, Y_{ij}, of pupil j in class i (j = 1,..., n persons per class; i = 1,..., k classes; for simplicity we assume equal-size classes) and the performance level X_{ij}, of the pupil prior to entering the class (i.e., the pretest or some measure of entering ability), then the relation of X_{ij} to Y_{ij} can be decomposed into between-class and within-class components (Burstein and Linn, 1977; Cronbach, 1976):

$$
\begin{aligned}
Y_{ij} - \bar{Y}_{..} &= \beta_b(\bar{X}_{i.} - \bar{X}_{..}) && \text{Predicted Between-Class} \\
&+ \bar{Y}_{i.} - \beta_b(\bar{X}_{i.} - \bar{X}_{..}) && \text{Adjusted Between-Class} \\
&+ \beta_w(X_{ij} - \bar{X}_{i.}) && \text{Pooled Within-Class Slope} \\
&+ (\beta_i - \beta_w)(X_{ij} - \bar{X}_{i.}) && \text{Specific Within-Class} \\
&+ \varepsilon_{ij} && \text{Specific Residual Associated with Person ij}
\end{aligned}
\qquad (1)
$$

In the above equation, β_b is the between-class slope from the regression of $\bar{Y}_{i.}$ on $\bar{X}_{i.}$, β_w is the pooled within-class slope from the regression of $(Y_{ij} - \bar{Y}_{i.})$ on $(X_{ij} - \bar{X}_{i.})$ <u>across all</u> classrooms, and the β_i are the <u>specific</u> within-class slopes from the regression of Y_{ij} on X_{ij} within the i classrooms.

The possible substantive interpretations of specific components and sets of components are important here. (See particularly descriptions of alternative analytical models and the section on slopes as indices). The key elements are the between-class slope, the adjusted between-class effect, the pooled within-class slope, and the specific within-class slopes.

Often, Equation (1) can be modified so that we have a global measure, T_j, of classes (e.g., class membership, teacher quality, or treatment-group) rather than the aggregation of individual scores represented by $\bar{X}_{j.}$. In what follows we shall refer to the effects associated with either $\bar{X}_{j.}$ or T_j as class effects without loss of generality.

One useful treatment of the multilevel analysis was provided by Cronbach (1976). A succinct statement of Cronbach's justification for his proposed analysis is that the usual overall between-student analysis combines two kinds of relationships--those operating between collectives (reflected in β_b and adjusted class effects) and those operating among persons within collectives (reflected in β_w and β_j)--into a composite that is rarely of substantive interest (see, E.g., Cronbach, 1976, pp. 10.3ff.). Cronbach reminds us that β_t, the overall between-student coefficient from the regression of Y_{ij} on X_{ij},

$$(2) \quad Y_{ij} - \bar{Y}_{..} = \beta_t(X_{ij} - \bar{X}_{..}) \quad ,$$

has been shown by Duncan, Cuzzort, and Duncan (1961, p. 66) to be a composite of β_b and β_w:

$$(3) \quad \beta_t = \eta_X^2\beta_b + (1 - \eta_X^2)\beta_w$$

where η_X^2 is the intraclass correlation or correlation ratio of X. He (1976; Cronbach and Webb, 1975) goes on to recommend that between-group effects and individuals-within-group effects should be examined separately.

In its most parsimonious form, Cronbach would examine the following:

$$(4) \quad \underline{\text{Between Groups}}: \quad \widehat{\bar{Y}_{j.} - \bar{Y}_{..}} = b_{\bar{Y}T}T + \beta_b(\bar{X}_{j.} - \bar{X}_{..})$$

where the $b_{\bar{Y}T}$ is the effect of teachers on mean outcomes after controlling for between-class differences in inputs.

(5) Pooled Within-Groups: $\widehat{Y_{ij} - \bar{Y}_{i.}} = \beta_w(X_{ij} - \bar{X}_{i.})$

Thus, Cronbach's primary concerns are with the adjusted collective effects of instruction as reflected by the adjusted class mean outcomes and with the overall redistributive properties of classroom instruction as reflected by the pooled within-class regression, β_w.

Empirical Results from Multilevels of IEA Data

For the time being, we focus on the two estimators of most interest to Cronbach, between-group regression coefficients and the corresponding pooled within-group coefficients. Recent empirical analyses of data from the IEA studies (Burstein, 1977c; Burstein, Fischer, and Miller, 1978) dramatically demonstrate the distinct differences in interpretation when one moves from a between-school to a within-school analysis. In an investigation of educational effects models for 14-year-olds from the U.S. and Sweden in the IEA science achievement study, Table 1, we found that the effects of family background on science achievement were substantial, as usual, in the between-schools analysis of U.S. data but much smaller in Sweden. In fact, for 14-year-olds, R^2_{total} was larger than $R^2_{between\text{-}schools}$ in Sweden, which would be atypical for analysis of U.S. data.

In contrast, the effects of family background in the pooled within-school analyses for the U.S. were substantially smaller and were essentially the same as the effects found in the within-school analysis for Sweden. One possible substantive explanation for these findings is that the two types of analyses reflect, on one hand, distinctions between the countries in the political order governing the distribution of pupil backgrounds and school resources (i.e., the predominance of local control and community determination of school resources in the U.S. vs.

national control and a policy of uniformity of resources for Sweden) and, on the other, the similarity between countries in the operation of the social order within schools (i.e., interpersonal allocations of rewards within an institution).

Insert Table 1

There are further substantive questions that the above example might address, but the methodological point is clear: different types of analyses of multilevel data address different questions and typically research on schooling asks questions at multiple levels.

Within-Group Slopes as Indices in Between-Group Analyses

Once it is determined that the questions of interest and/or statistical considerations warrant some analyses of aggregated data, the types of between-group effects one expects to find remain to be specified. In particular, when one's purpose is to determine factors affecting pupil performance, it is possible that the analyses of between-group (class, school, etc.) means can hide important differences in the within-group distribution of pupil outcomes and educational inputs.

Several aspects of current schooling practices lead us to expect that within-school and within-class distributions of pupil performance vary. First, schools (classes) do differ in the distribution of educational performance as well as in mean performance. Moreover, schools with the same mean outcome often exhibit different distributions of

performance within school. An analysis of means alone could not be expected to account for such distributional differences.

Second, a variety of educational theories about the effects of specific schooling practices on within-group behavior argue for an examination of distributional properties other than group means. Obviously, at least the variability of performance is of interest in studies comparing individualized, competency-based, or open educational instructional programs with more traditional instructional practices. Also, research on the interaction between teaching style and learning style would lead one to expect variability of outcomes for pupils with similar entering characteristics and preferences taught by teachers with differing instructional styles.

Finally, the idea of using distributional characteristics in addition to the mean as criterion measures has been shown previously to merit consideration (Lohnes, 1972; Klitgaard, 1975; Brown and Saks, 1975a, 1975b). Lohnes (1972) found that standard deviations and skewness indices added to the explanatory power of means in his analyses of data from the Cooperative Reading Project. Klitgaard (1975) and Brown and Saks (1975a, 1975b) found that school and school district standard deviations exhibited more significant relations with school characteristics than did school and school district means.

Though they sought answers to different questions and used different methodologies, Lohnes, Brown and Saks, and Klitgaard apparently share our belief that educational outcomes are multifaceted and incompletely measured by simple group averages. There also seems to be consensus that educational theory can be developed which will link pupil entering characteristics and characteristics of the educative process to distributional properties of educational outcomes.

CSE
ARCHIVE COPY

We (Burstein, 1976; Burstein and Linn, 1976, 1977) have elaborated a theory for the use of within-group slopes of outcomes on inputs as criterion in educational effects studies. Wiley (1970) may have been the first to suggest this strategy. Also, Karweit and Fennessey (1977) have independently suggested that differences in slopes across schools deserve further consideration.

Our justification for considering within-group slopes as outcomes derives much of its impetus from research on aptitude-treatment interactions (Snow, 1977) and from evidence of slope differences among colleges (Rock, Baird, and Linn, 1972). In its simplest form, we expect that different combinations of teachers and instructional practices will result in varying distributions of educational outcomes for pupils with similar entering characteristics. For example, it might be hypothesized that there are teachers who are equally effective in obtaining mean performance, but yield varying slopes because some teachers use compensatory instructional practices which emphasize the improved performance of lower-ability students while others allow each child to learn at his/her own rate. (We would expect a flatter slope in the former case than in the latter.)

Burstein and Linn (1976, 1977) compared alternative analytical models for identifying educational effects for sets of hypothetical classrooms with heterogeneous slopes. The key findings were that, for the conditions studied, heterogeneous within-class slopes were shown to make important differences in identified effects, ones which were not swamped by sampling variability in the estimation of slopes, and certain analytical strategies exhibited good properties even in the presence of heterogeneity.

Although within-group slopes are conceptually appealing indices of educational effects, three points warrant further examination. First, it must be determined that slopes are sufficiently stable. Second, it

must be demonstrated that slopes are potentially distinct from other group indices (e.g., pre and posttest means and standard deviations) in realistic situations. Finally, there have to be realistic cases in which slopes are related to school and class characteristics after controlling for other background measures and other indices of group outcomes. We have already begun to investigate these points (see below) and can describe how we plan to proceed.

Stability of Slopes

The sampling variability of within-group slopes is substantially greater than that of the mean. For small samples, e.g., the size of a classroom, the sampling error of a slope is so large that it is questionable whether real differences in slopes may reasonably be distinguished from the noise; moreover, any outlier can dominate the slope. If the real differences in slopes are as large as those generated in Burstein and Linn (1976, 1977), then it is important to take them into account. Whether the differences in real classrooms are of similar magnitude is somewhat problematic at this stage, however.

Since students within a classroom are not a random sample, but possibly are better thought of as fixed once the classroom is chosen, it is not clear how best to investigate the relative magnitude of signal and noise in the differences among within-classroom slopes. Linn and Burstein (1977) found little support for the notion that slopes varied systematically when a posttest in reading was regressed on a pretest in reading, and only limited support for the notion based on a similar set of regressions for math using small samples of classrooms from the ETS BTES study (McDonald and Elias, 1976). (See Figures 1 and 2.) But

these analyses were based on traditional confidence intervals which treated each class as if the students in it were a random sample from a population. As already noted, the random sampling model is questionable in this situation.

Insert Figures 1 and 2 about here

While random sampling of students may not provide the best model, there is a need to allow for disturbances in the observed slope due to idiosyncratic occurrences at the time of measurement. Just as an individual's observed score is distinguished from an underlying true score in classical test theory, there is a need to distinguish between the observed measure for the group (in this case the slope) and an underlying "true" slope.

Several approaches can be used to investigate the relative size of signal and noise in the within-group slope estimates. Confidence intervals can be computed for the within-group slopes for selected sets as was done by Linn and Burstein (1977) for BTES data. The Jacknife procedure can also be used to estimate slopes and confidence.

Relations of Slopes to Other Group-Level Indices

If slopes are to provide a useful addition to the array of outcomes, they must be distinct from other indices. Burstein and Linn (1977) and Burstein (1977c) have investigated this property of slopes. For three separate data sets (BTES data on classrooms collected by ETS (McDonald and Elias, 1976); Michigan Assessment Data on schools reported in Marco (1974); and IEA data on schools reported in Burstein (1977c),

(see Tables 2 through 4), they found that though pretest and posttest means correlated with each other in the range of .7-.9 and pretest and posttest standard deviations correlated in the range of .5-.8, the correlation of within-group slope with either means or standard deviation (or, for that matter, with skewness and kurtosis indices and sample size) are much lower and, except for the pretest standard deviations (which are spuriously related to slopes), are rarely significant.

Insert Tables 2, 3, and 4 about here

The results cited suggest that slopes are sufficiently distinct from means and standard deviations to warrant further consideration.

Relation of Slopes to School and Class Characteristics

The final line in an investigation of the potential utility of within-group slopes involves their relationships to measures of school and classroom processes. Preliminary results of an analysis of science achievement data on U.S. 14-year-olds in the IEA study (Burstein, 1977c) provided tantalizing evidence of the possible payoff from this activity. Burstein found that the within-school slopes of science achievement on a verbal ability measure (assessed concurrently) were significantly and positively related to school mean responses of pupils on indices of exposure to science instruction and of the degree to which pupils reported instructional practices which emphasized exploration (discovery methods of instruction). (See Tables 5 through 8.) These significant results occurred despite controls for pretest and posttest means and standard deviations and pupil home background measures.

Insert Tables 5 through 8 about here

The results described above fit in well with other recent research on informal/open/individually-guided/unstructured instruction (see particularly Rosenshine (1978) and Stebbins, et al. (1977)). Instruction which emphasizes student self-direction (selection) of learning goals and methods tends to exacerbate pre-existing differences in pupil skills. Higher-ability students tend to make more appropriate choices and achieve at a faster rate than lower-ability students.

The steeper within-group slopes with greater opportunities for exposure to instruction and with greater emphasis on individual exploration cited above are consistent with expectations from other research and suggest the need for similar investigations with other data sets.

Estimating Within-Group Dependency in Multilevel Analysis

The Problem of Dependency among Observations within Units

The problem of dependence among observations within groups is endemic to research on hierarchically nested school data, and can be especially critical when intact classrooms are investigated. Cronbach and Webb (Cronbach, 1976; Cronbach and Webb, 1975; Webb, 1977) have argued that when intact groups are assigned to instructional treatments, the students in those treatments cannot be considered independent units and therefore, the typical analyses based on all individuals pooled across groups can be justifiably criticized as inappropriate.

The crucial problem in ignoring group membership is that educational treatments are not administered independently to individuals (Wiley, 1970). Individuals within the classroom have shared experiences. This nonindependence of individuals within the group can be expressed by an intraclass correlation structure. The consequences of ignoring this intraclass structure (i.e., treating individuals as independent by ignoring group membership) are serious (Walsh, 1947; Weibull, 1953).

Recent work by Glendening and Porter (Glendening, 1976a, 1976b; Glendening and Porter, 1975, 1976) provides a thorough discussion of the problem in the experimental design frame of reference. The work of Glass and Stanley (1970), Peckham et al. (1969) and Raths (1967) is also summarized by Glendening and Porter.

Glendening (1976a, 1976b) simulated the effects of violating the assumption of independence within the context of a balanced, two-level hierarchically-nested design, with subjects (S) nested within classrooms (C) and classrooms nested within treatments (T). She operationally defined independence as that condition wherein the expected mean square between classrooms (EMS(C:T), equals that within classrooms, EMS(S:CT). She found that a model with the pupil as the unit or a conditional model where a preliminary test of independence is followed by a choice of unit of analysis for testing treatment effects, yielded spuriously small error terms and therefore, too liberal tests of treatment effects. Glendening concluded that the researcher must choose a priori between the class (dependence) or student (independence) as the unit, but acknowledged the complications of obtaining prior knowledge about independence of response.

While Glendening and Porter focused on the implications of intra-calss correlation for the analysis of experimental data, Webb (1977) was concerned with the antecedents of such intraclass relations in research on group process. Webb compared learning in interacting groups and leraning singly, attempting to explain differences as a function of the characteristics of the individual, the group, and the group process. The group process results provided a key to understanding why some students learned best in interacting groups, whereas others did best learning singly. In general, group members who actively participated in discussions did better than those who did not actively participate, and did at least as well as after individual learning. Whether a pupil actively participated was related to the pupil's ability ranking within the group and the range of ability in the group. Knowing the abilities of the students in a group, one could predict fairly well who interacted with whom and, consequently, who did best.

The results of this highly structured study suggest that knowledge of group processes in a particular class is crucial for understanding the degree to which students are working together--and therefore crucial for estimating degree of dependence in the class. Studying group process may be the only way to get at this dependence. Unless students in a class are receiving completely individualized instruction, rarely will it be tenable to base analyses on an intraclass correlation of zero. Unless all students are receiving exactly the same instruction and interact with fellow students in the same manner and amounts of time, an intraclass correlation of one is unreasonable. Examination of lower-level processes will help locate the intraclass correlation on the continuum between 0 and 1.

Webb suggests that the above procedures may be generalized to real teacher-taught classrooms, considering interactions between teacher and students, interactions among students and characteristics of students (abilities, personality variables) and teachers. In the long run, one hopes to be able to predict student performance from a combination of these variables.

Clearly, research on most educational phenomena will involve dependent observations. Moreover, dependence cannot be viewed as an all-or-none phenomenon--it is a matter of degree. It depends on what is being measured (the outcome) and the "treatments" or "causes" under study. It is also a function of the composition of the units and the nature of the grouping mechanism as Webb (1977) has demonstrated. Therefore tests for independence and adjustments for intraclass correlations are more appealing than automatic aggregation to the classroom level.

Analytical methods are needed which will account for the degree of dependency and make adjustments, where appropriate, to the estimated effects and associated estimates of precision. Moreover, estimators of dependency may be useful as indicators of classroom process. That is, it may be possible to relate these estimated relationships to characteristics of students, teachers, and instructional context.

Concluding Remarks

The topics discussed in this paper are a subset of a broader range of issues and problems which require more attention over the next few years. Table 9 lists a variety of types of studies and types of outcomes for which multilevel analysis issues must be resolved.

It is unclear what form the final products of the investigation of the analysis of multilevel data will take, but it is possible to imagine

the following scenario. As a preamble, we point to a trend developing in educational evaluation for the conduct of what Glass (1976) has termed "meta-analyses" (see alse Light and Smith, 1972). Persons conducting meta-analyses seek to accumulate knowledge about the impact and characteristics of a particular educational innovation by aggregating findings across numerous investigations of the phenomena.

There would seem to be a natural parallel to meta-analysis which is relevant to the examination of alternative methodological approaches for the analysis of multilevel data. There are two key obstacles to the development of appropriate methodologies in this context. First, the available methodological approaches vary greatly in the degree to which they are theory-based as opposed to ad hoc. Second, all currently available empirical data sets suffer from a variety of inadequacies which, taken singly, limit their utility for comparing alternative methodological approaches.

We believe that it is important to identify approaches which are practically viable as well as theoretically sound and which are usable with actual as well as hypothetical data. Therefore, we propose that in addition to the studies of the variation in analytical properties across approaches with hypothetical data, the alternative approaches should be applied to a wide variety and sizeable number of actual data sets, each with a potentially differing set of inadequacies. In this way we hope to learn more about both the methods (e.g., which are more generally usable; which behave similarly for specific kinds of data sets) and the influence of data limitations on methods (e.g., the exclusion of what types of information makes different approaches impractical or unattractive).

Table 1. Between-Student, Between-School and Pooled Within-School Analyses of Factors Affecting Science Achievement (RSCI): 14 year olds from the IEA Study.

UNSTANDARDIZED

	UNITED STATES			SWEDEN		
	Pupil	School	Within School	Pupil	School	Within School
RWK	.842 (22.62)	.870 (6.33)	.799 (21.12)	.724 (17.18)	.489 (2.68)	.736 (17.34)
GRADE	1.924 (5.34)	1.402 (1.69)	2.321 (5.33)	2.984 (7.3[illegible]	4.442 (2.60)	2.893 (7.1[illegible]
SEX	-4.168 (11.95)	-6.633 (3.87)	-3.870 (11.37)	-5.158 (13.73)	-4.035 (1.99)	-5.275 (14.27)
SCISTUDY	.006 (4.20)	.011 (2.36)	.006 (3.68)	-.012 (6.01)	-.011 (1.72)	-.012 (5.49)
EXPLORE	.136 (2.63)	.238 (1.47)	.077 (1.42)	-.109 (1.43)	.597 (1.66)	-.141 (1.85)
AUTHITY	.041 (.76)	-.096 (.69)	-.024 (.40)	.188 (2.26)	.326 (.96)	.169 (2.00)
POPSOC	.435 (5.66)	.798 (2.36)	.286 (3.64)	.232 (2.96)	.336 (1.13)	.270 (3.32)
MOMED	1.023 (4.02)	.328 (.27)	.838 (3.31)	.735 (3.73)	-1.429 (1.72)	.944 (4.73)
BOKHOM	1.487 (6.35)	3.545 (3.23)	1.101 (4.78)	1.205 (4.54)	1.387 (1.04)	1.205 (4.59)
R^2	.396	.726	.317	.344	.339	.354

STANDARDIZED

RWK	.444	.434	.428	.363	.270	.365
GRADE	.101	.099	.103	.159	.251	.153
SEX	-.219	-.213	-.223	-.275	-.191	-.284
SCISTUDY	.078	.138	.072	-.131	-.170	-.120
EXPLORE	.049	.[illegible]3	.028	.046	.159	-.038
AUTHITY	-.014	-.037	-.006	-.029	.129	.041
POPSOC	.110	.200	.073	.061	.095	.067
MOMED	.078	.022	.066	-.076	-.176	.096
BOKHOM	.125	.250	.097	.094	.107	.094

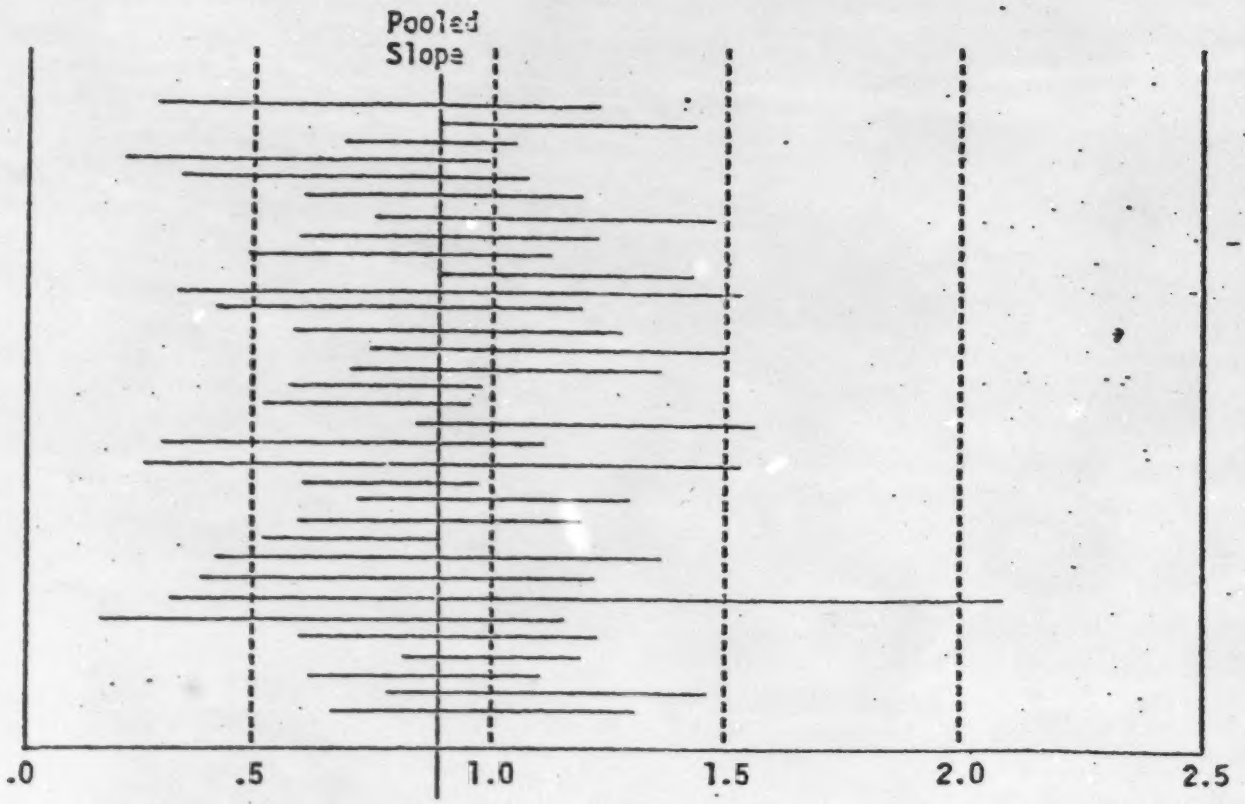

FIGURE 1

Ninety Percent Confidence Levels for Regressions of Posttest in Reading on Pretest in Reading for 33 Second-Grade Classrooms (Data from McDonald & Elias, 1976)

Source: Linn & Burstein, 1977

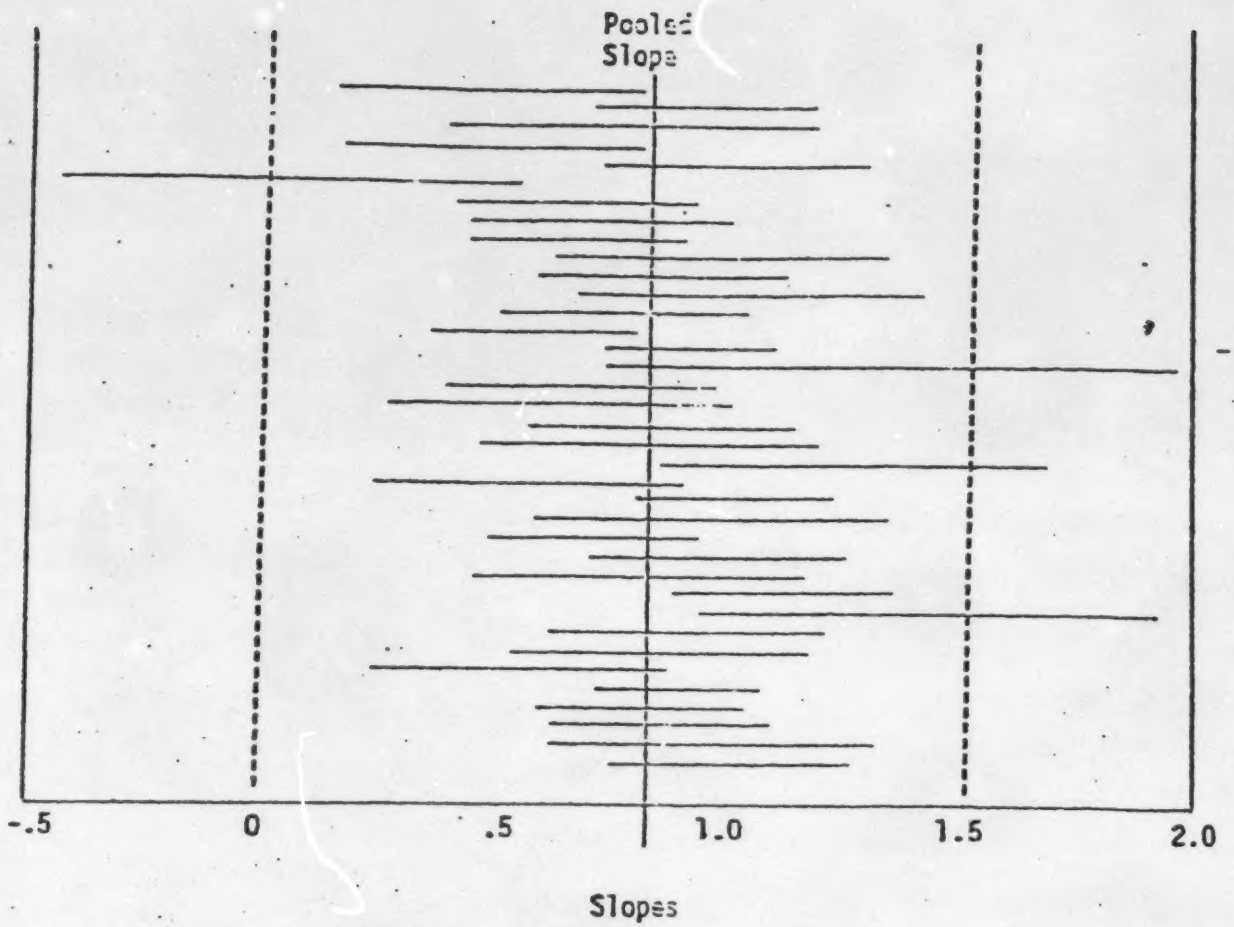

FIGURE 2

Ninety Percent Confidence Intervals for Regressions of Posttest in Mathematics on Pretest in Mathematics for 37 Second-Grade Classrooms (Data from McDonald & Elias, 1976)

Source: Linn & Burstein, 1977

Table 2. Correlations from the ETS BTES Data. (N = 39; McDonald, et al., 1976)

		MATH						
		B	$\bar{Y}$	$\bar{X}$	S_y	S_x	N	Mean
READING	B		.10	.14	.59*	.13	.29*	.85
	$\bar{Y}$	-.02		.88*	-.09	-.24	.27	177.74
	$\bar{X}$	.09	.93*		-.12	-.34*	.20	152.50
	S_y	.48*	-.18	-.21		.78*	-.38*	18.31
	S_x	-.04	.12	.02	.65*		-.31*	18.41
	N	-.35*	.23	.19	-.19	-.05		
	Mean	.90	173.20	156.27	34.03	31.78		

Source: Linn & Burstein, 1977

Table 3. Marco Michigan Assessment Data (N = 76)

		WORD KNOWLEDGE							
		B	$\bar{Y}$	$\bar{X}$	S_y	S_x	N	Mean	S.D.
READING	B		-.22*	-.18	.16	-.41*	-.16	.81	.18
	$\bar{Y}$	-.13		.86*	.13	.09	-.09	60.63	5.01
	$\bar{X}$	-.04	.89*		.11	.15	-.07	53.81	4.25
	S_y	.49*	-.04	-.09		.71*	.05	10.09	1.47
	S_x	-.30*	.11	.09	.49*		.07	9.38	1.91
	N	-.12	-.05	-.01	-.01	.14			
	Mean	.69	51.18	58.13	10.44	10.65			
	S.D.	.16	4.74	4.83	1.80	1.71			

Source: Burstein (1977c); Marco (1974)

Table 4. IEA U.S. Data Correlations among Descriptive Statistics (N = 107)

	B	$\bar{Y}$	$\bar{X}$	S_y	S_x
$\bar{Y}$	.35*				
$\bar{X}$	.35*	.77*			
S_y	.58*	.60*	.40*		
S_x	-.11	.19*	-.15	.34*	
Mean	.90	33.47	27.54	8.08	4.45
S.D.	.45	5.00	2.41	2.05	1.00

Y = Raw Science Total

X = Raw Word Knowledge

Source: Burstein (1977c)

Table 5. Variables included in the regression analyses.

RSCI = Raw total science score

RWK = Raw word knowledge score

SEX = Sex of student (1=male, 2=female)

POPOCC = Ascending cale of student's report of father's occupation (1=unskilled, manual, . . . 9=professional, technical requiring higher education)

BOKHOM = Ascending scale of student's report of number of books in the home (1=none, 2=1-10, 3=11-25, 4=26-50, 5=51 or more)

GRADE = Grade in school (5 to 11)

SCISTUDY = Ascending scale of student's report of exposure to science based on formula

$$\text{SCISTUDY} = \frac{\text{INTENSITY} + (\text{YRSSTUDY} * \text{INTENSITY})}{\text{YRSSTUDY}}$$

where YRSSTUDY = 724 + 725 + 726 + 727

INTENSITY = (732 + 733 + 734 + 735 + 736 + 737 + 738 + 739)/169

169 = number of weeks per year (19-50)

724-727 = Years of study of General Science, Biology, Chemistry Physics (1-5; never studies, 1 year; 1-3 years, 3-5 years, greater than or equal to 5)

732-735 = Hours of homework per week of General Science, Biology, Chemistry, Physics (1-5; did not take, less than or equal to 1 hour, 1-3 hours, 3-5 hours, greater than 5 hours)

736-739 = Hours of instruction per week (1-5; did not take, less than 2 hours, 2-4 hours, greater than 6 hours)

EXPLORE = Ascending scale of student's report of degree of use of exploration in science in science instruction, sum of scale responses on the following items (reversed items indicated by (-))

Science mostly learned through experiments -- (+)
Science reading encouraged -- (+)
Lab reports for homework -- (+)
Science theory more than practice -- (-)
Fieldwork encouraged -- (+)
Participation in Labs -- (+)
Experiments in science lessons -- (+)
Labs, student problems, teacher helps solve -- (+)
Labs, devise own experiments -- (+)
Labs, teacher problems, student solves -- (+)
Labs, teacher questions on experiments -- (+)

Table 6. School-Level Regressions of Means, Standard Deviations, and Slopes on Background and School Characteristics: Population II, IEA, U.S., Unstandardized Regression. (N = 107 schools)

Independent Variable	Dependent Variable[a]		
	$\overline{RSCI}$	SD(RSCI)	B(RSCI,RWK)
RSCI	--	.162* (2.33)	-.018 (1.19)
RWK	1.102* (8.48)	.052 (.53)	.040 (1.65)
SD(RSCI)	--	--	.145* (6.87)
SD(RWK)	1.008* (4.09)	.441* (2.40)	-.128* (3.25)
SEX	-6.846* (4.69)	-.900 (.81)	-4.20* (1.81)
Home Background			
POPSOC	.686* (2.64)	-.012 (.06)	-.001 (.03)
BOKHOM	2.848* (3.06)	.811 (1.21)	-.088 (.63)
School Characteristics			
GRADE	1.230 (1.62)	-.839 (1.58)	-.017 (.15)
SCISTUDY	.125* (2.98)	.012 (.39)	.016* (2.58)
EXPLORE	.222 (1.47)	.144 (1.38)	.055* (2.53)
R^2	.81	.46	.53

[a] t-statistics in parentheses.

Table 7. School-Level Regressions of Means, Standard Deviations, and Slopes on Background and School Characteristics: Population II, IEA, U.S., Standardized Coefficients. (N = 107 schools)

Independent Variable	Dependent Variable		
	RSCI	SD(RSCI)	B(PSCI,RWK)
RSCI		.395	-.194
RWK	.531		.215
SD(RSCI)			.656
SD(RWK)	.201	.214	-.281
SEX	-.216		-.146
Home Background			
POPSOC	.171		
BOKHOME	.208	.144	
School Characteristics			
GRADE		-.136	
SCISTUDY	.147		.209
EXPLORE		.112	.196

Table 8. School-Level Regressions of Means, Standard Deviation, and Slopes on Background and School Characteristics: Population II, IEA, U.S., Unstandardized Regression. (N = 107 schools) (Excluding means and standard deviations from regressions)

Independent Variable	Dependent Variables[a] RSCI	SD(SCI)	B(RSCI,RWK)	
SD(SCI)	--	--	.121* (33.54)	--
SEX	-8.409* (4.42)	-2.387* (2.20)	-.371 (1.60)	-.611* (2.53)
Home Background				
POPSOC	1.239* (3.75)	.219 (1.17)	.003 (.08)	.030 (.66)
BOKHOM	5.699* (4.99)	2.058* (3.16)	-.104 (.73)	.145 (.93)
School Characteristics				
GRADE	2.912* (3.09)	-.459 (.86)	.064 (.56)	.008 (.06)
SCISTUDY	.144* (2.66)	.029 (.92)	.018* (2.79)	.022* (2.90)
EXPLORE	.357* (1.84)	.255* (2.30)	.041* (1.73)	.071* (2.70)
R^2	.66	.34	.41	.21

[a] t-statistics in parentheses.

Table 9. Classifications of types of studies and types of outcomes for the investigation of educational effects.

I. TYPE OF STUDY

A. MANIPULATION

1. EXPERIMENTAL/TRUE -- "Units" assigned to alternative treatments or Treatment/Non-treatments; some form of manipulation

a. Random Assignment of Pupils from Classrooms to Treatments -- Pupils randomly assigned to treatment conditions; treatment outside of normal class routine; treatment non-group work

b. Random Assignment of Pupils from Classrooms to Groups -- Pupils randomly assigned to treatment groups; treatment outside normal class routine

c. Random Assignment of Pupils to Classes -- Pupils randomly assigned to classes; classes randomly assigned to treatments

d. Random Assignment of Partial Classes to Treatments -- Portions of class randomly assigned to different treatment conditions.

e. Random Assignment of Intact Classes to Treatments -- Students assigned to classes on unknown non-random basis; intact classes assigned to treatments

2. EXPERIMENTAL/ATI -- Conditions under I with additional question of interaction with entering characteristics

3. EXPERIMENTAL/LONGITUDINAL -- Repeated measurement (mastery testing, sequential analysis of behavior and interaction patterns, persistence) in context of empirical studies

B. NON-MANIPULATION

1. NON-EXPERIMENTAL/CROSS-SECTIONAL -- Large-scale cross-sectional survey of pupils, teachers/classrooms, schools, etc. for purpose of establishing educational school/teacher effects modeld (e.g., Coleman, IEA)

2. NON-EXPERIMENTAL/LONGITUDINAL -- Large-scale longitudinal survey (e.g., income maintenance, voucher study, Follow Through Evaluation, BTES?)

3. NON-EXPERIMENTAL/OUTLIER (RESIDUAL) ANALYSIS -- Develop indices of effects of system over and beyond what can be anticipated by entering characteristics (Klitgaard, Kiesling, Rassmussen, Marco, Convey, New York, California)

4. NON-EXPERIMENTAL/CONTEXTUAL (COMPOSITIONAL) EFFECT -- Same as IV but question is whether the composition/frog-pond/normative climate of institution has an effect.

Table 9. continued

II. TYPE OF OUTCOME

A. SHORT TERM -- Duration of a lesson to, say, a year

1. Specific Cognitive Objective -- Single content domain/objective in an instructional sequence

2. General Cognitive Objective -- Standardized achievement test or total score over multiple objectives CRM

3. Affective Objective -- Attitude toward self and subject matter, efficiency

4. Group Behavior -- Peer socialization, group cohesiveness, group interaction

B. LONG TERM -- Duration of multiple years, retrospective academic antecedents

1. General Cognitive Outcome -- Standardized test or cumulative grades (e.g., SAT as outcome prediction of future grades from earlier test scores)

2. Educational Attainment -- Level of education

3. Occupational Attainment -- Level of occupation (social stratification theory)

4. Career Plans/Career Satisfaction

5. General Mental Health

REFERENCES

Alwin, D.F. "Structural Equations Models of Contextual Effects." Technical Report No.20, Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisconsin: Vasquez Associates, Ltd., October, 1976.

Averch, H., S.J. Caroll, T.S. Donaldson, H.J. Kiesling and J. Pincus How Effective is Schooling?, The Rand Corporation, R-956-PCSF/RC, 1972.

Berliner D.C., L.S. Cahen, N. Filby, C. Fisher, R. Marliave, J. Moore "Proposal for Phase III B," Beginning Teacher Evaluation Study July 1, 1976 - June 30, 1978. Far West Laboratory for Educational Research and Development, 1976.

Bock, G. L.B. Stebbins and M.S. Stearns Interim Evaluation of the National Follow Through Program 1969-1971: A Technical Report, Menlo Park, CA.: Stanford Research Institute, February, 1973.

Box, G.E.P. "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems. I. Effect of Inequality of Variance in the One-way Classification. Annals of Mathematical Statistics, 1954, 25: 290-302.

Brophy, J.E. "The Student as the Unit of Analysis," Research Report No. 75-12, University of Texas at Austin, 1975.

Brown, W. and D.H. Saks "The Production and Distribution of Cognitive Skills within Schools." Journal of Political Economy, 1975, 83: 571-593.

Brown W. and D.H. Saks "Proper Data Aggregation for Economic Analysis of School Effectiveness," Public Use Data, 1975, 3: 13-18.

Burks, B. "Statistical Hazards in Nature-Nurture Investigations." 27th Yearbook of the National Society for the Study of Education, Part 1, 1928: 9-33.

Burstein, L. "The Choice of Unit of Analysis in the Investigation of School Effects: The IEA in New Zealand," New Zealand Journal of Educational Studies, 11 (May): 11-24, 1976.

Burstein, L. and I.D. Smith "Choosing the Appropriate Unit for Investigating School Effects," The Australian Journal of Education, 21, 1: 65-79, 1977.

Burstein, L."Assessing the Differences of Between-Groups and Individual Regression Coefficients." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, Ca. April, 1976.

Burstein, L. "The Use of Data from Groups for Inferences about Individuals in Educational Research." Unpublished Doctoral Dissertation, Stanford University, December, 1975. (Also, Technical Report No. 7, Consortium on Methodology in Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez, Associates, Ltd., December, 1975.

Burstein, L. "Data Aggregation in Educational Research: Applications," Technical Report No. 1, Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez Associates, Ltd., March, 1975 (Also, paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC, April, 1975.)

Burstein, L. "Issues Concerning Inferences from Grouped Observations," Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April, 1974.

Burstein, L. "Alternative Approaches for Assessing Differences Between Grouped and Individual-Level Regression Coefficients," 1977, article in review.

Burstein, L. and M.T. Hannan, Progress Report NIE-C-74-0123: Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee Wisc.: Vasquez Associates, Ltd., July, 1975.

Burstein L. and M.T. Hannan Second semiannual progress report on the Consortium on Methodology for Aggregating Data in Educational Research, Report to the National Institute of Education under Grant NIE-C-74-0123 to Vasquez Associates, Ltd., Milwaukee, Wisc., July 1975.

Burstein, L. and M.T. Hannan First semiannual progress report on the Consortium on Methodology for Aggregating Data in Educational Research, Report to the National Institute of Education under Grant NIE-C-74-0123 to Vasquez Associates, Ltd., Milwaukee, Wisc., January 1975.

Burstein, L. and M.T. Hannan A Proposal for a Consortium on Methodology for Aggregating Data in Educational Research, Submitted by Vasquez Associates, Ltd., Milwaukee, Wisc. to the National Institute of Education, May 1974.

Burstein, L. and T.R. Knapp "The Unit of Analysis in Educational Research," Technical Report No. 4, Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez Associates, Ltd., July, 1975.

Burstein L. and R.L. Linn "Detecting the Effects of Education in the Analysis of Multilevel Data: The Problem of Heterogeneous Within-Class Regressions," Paper presented at the Conference on Methodology for Aggregating Data in Educational Research, Stanford, CA, October, 1976

Burstein, L. and R.L. Linn "Alternative Analytical Models for Identifying Educational Effects in the Presence of Heterogeneous Within-Group Relations of Input to Outcome," 1977, article in review.

Cline, M.D., et al. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Vols. 1A; 1B. Cambridge, Mass.: Abt Associates, 1974.

Cochran, W.G. "Some Consequences when the Assumptions for the Analysis of Variance are Satisfied, Biometrics, 1947, 3: 22-28

Cramer, J.S. "Efficient Grouping, Regression and Correlation in Engle Curve Analysis." Journal of American Statistical Association, 1964: 233-249.

Cronbach, L.J., D.R. Rogosa, R.E. Floden and G.G. Price "Analysis of Covariance in Nonrandomized Experiments: Parameters Affecting Bias," Occasional Paper, Stanford Evaluation Consortium, August, 1977.

Cronbach, L.J. with assistance of J.E. Deken and N. Webb "Research on Classrooms and Schools: Formulation of Questions, Design, and Analysis," Occasional Paper, Stanford Evaluation Consortium, July, 1976.

Cronbach, L.J. and N. Webb "Between-Class and Within-Class Effects in a Reported Aptitude X Treatment Interaction: Reanalysis of a Study by G.L. Anderson." Journal of Educational Psychology, 1975, 67: 717-724.

Emrick, J.A., P.H. Sorenson, and M.S. Stearns Interim Evaluation of the National Follow Through Program 1969-1971. A Technical Report. Menlo Park, CA.: Stanford Research Institute, February, 1973.

Feige, E.L. and H.W. Watts "An Investigation of the Consequences of Partial Aggregation of Micro-Economic Data." Econometrica, 1972, 40: 343-360.

Glass, G.V. and J.C. Stanley Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice Hall, 1970.

Glendening, L. "The Effects of Correlated Units of Analysis: Choosing the Appropriate Unit." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976. (Also, unpublished Doctoral Dissertation, Michigan State University, East Lansing Michigan, 1975.)

Glendening, L. and A.C. Porter "Interdependence and Selecting the Appropriate Unit of Analysis (or the Appropriate Analytical Model)." Technical Report No. 23, Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez Associates, Ltd., October, 1976.

Glendening L. and A.C. Porter "The Effects of Correlated Units of Analysis: Violating the Assumption of Independence." A Working Paper: Measurement and Methodology Program, NIE, November, 1974.

Goodman, L.A. "Some Alternatives to Ecological Correlation." American Journal of Sociology, 1959, 64: 610-625.

Haney, W. "Units of Analysis Issues in the Evaluation of Project Follow Through." Unpublished Report. Cambridge, Mass.: Huron Institute, 1974.

Hannan, M.T. "Estimation of Pooled Cross-Section and Time Series Models: Preliminary Monte Carlo Results." Proceedings of Conference on Policy Research: Methods and Implications, University of Wisconsin.

Hannan, M.T. "Aggregation Gain Reconsidered." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.

Hannan, M.T. Aggregation and Disaggregation in Sociology. Lexington, Mass.: Heath-Lexington, 1971.

Hannan, M.T. "Problems of Aggregation," in H.M. Blalock (Ed.) Causal Models in the Social Sciences. Chicago, Ill.: Aldine, 1971.

Hannan, M.T. Problems of Aggregation and Disaggregation in Sociological Research. Chapel Hill: Institute for Research in the Social Sciences, 1970.

Hannan, M.T. and L. Burstein "Estimation from Grouped Observations." American Sociological Review, 1974, 39: 374-92.

Hannan, M.T. and J. Freeman "Growth and Decline Processes in Organizations." American Sociological Review, 1975, 40: 215-28.

Hannan, M.T., J. Freeman, and J.W. Meyer "Specification of Models of Organizational Effectiveness." American Sociological Review, 1976, 41: 136-43.

Hannan, M.T., R. Rubinson, and J. Tuttle "The Causal Approach to Measurement Error in Panel Analysis: Some Further Contingencies," in H.M. Blalock (ed.), Social Science Measurement. Chicago: Aldine, 1974.

Hannan, M.T. and A. Young "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," in D. Heise (ed.) Sociological Methodology 1976-77. San Francisco: Jossey-Bass, 1977

Hannan, M.T. and A. Young "On Certain Similarities in the Analysis of Multi-Wave Panels and Multilevel Cross-Sections." Technical Report No. 16, Consortium on Methodology for Aggregating Data in Educational Research, 1975, Milwa kee, Wisc.: Vasquez Associates.

Hannan, M.T. and A. Young "Small Sample Results on Estimation from Grouped Observations." Technical Report No.24, Consortium on Method for Aggregating Data in Educational Research. Milwa kee, Wisc.: Vasquez Associates, Ltd., 1976.

Hannan, M.T., A. Young and F. Nielsen "Specification Bias Analysis of the Effects Grouping of Observations in Multiple Regression Models." Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April, 1975. (Also, Technical Report No. 2, Consortium on Methodology for Aggregating Data in Educational R search, Milwa kee, Wisc.: Vasquez Associates, 1975.)

Hauser, R.M. "Contextual Analysis Revisited." Sociological Methods and Research, 1974, 2: 365-375.

House, E.R., G.V. Glass, L.D. McCean, and D.F. Walker "No Simple Answer: Critique of the 'Follow Through' Evaluation," Report funded by the Ford Foundation, Semptember, 1977.

Keesling, J.W. "Components of Variance Models in Multilevel Analysis." Technical Report No. 18, Consortium on Methodology for Aggregating Data in Educational Research, Milwaukee, Misc.: Vasquez Associates, Ltd., 1976. (Also presented at Conference on Methodology for Aggregating Data in Educational Research, Stanford, Ca., October, 1976.)

Keesling, J.W. "Quantity of Schooling and Achievement: Some Ambiguities." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.

Keesling, J.W. "A Problem in the Aggregation of Student Data to the Level of School." Paper presented at the Annual Meeting of the American Educational Research Association, February, 1973.

Keesling,J.W. and D.E. Wiley "Regression Models for Hierarchical Data," Paper presented at the Annual Meeting of the Psychometric Society, Stanford University, 1974.

Keesling, J.W. and D.E. Wiley "Some Problems in the Application of Cross-Lagged Panel Correlations," Read at the 1969 Meeting of the American Educational Research Association.

Klitgaard, R.E. "Going Beyond the Mean in Educational Evaluation," Public Policy, 1975, 23: 59-79.

Lohnes, P. "Statistical Descriptors of School Classes," American Educational Research Journal, 1972, 9: 574-556.

Marliave, R., C. Fisher, N. Filby, M. Dishaw "The Development of Instrumentation for a Field Study of Teaching," Technical Report I-5, Beginning Teacher Evaluation Study, February, 1977.

McBee, J.K. and J.C. Fortune "Use of Distribution Parameters of Achievement Scores for Teacher Placement," Paper presented at the Annual Meeting of the American Psychologist Association, San Francisco, 1977.

McBee, J.K., J.C. Fortune, D.E. Hinkle and D.C. Berliner and J. Moore "The Use of Distribution Statistics to Evaluate the Effectiveness of Teachers in Unit of Instruction." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.

Mitchell, E.J., "Some Econometrics of the Huk Rebellion," American Political Science Review, 1969, 63: 1159-1173.

Nerlove, M. "Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross-sections," Econometrica, 1971, 39: 359-382.

Peckham, P.D., G.V. Glass, and K.D. Hopkins "The Experimental Unit in Statistical Analysis," Journal of Special Education, 3: 337-349, 1969.

Porter, A.C. "Some Design and Analysis Concerns for Quasi-experiments such as Follow Through," Paper presented at the 1972 American Psychologist Association Meeting, 1972.

Poyner, H. "Selecting Units of Analysis," in G.D. Borich (ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N.J.: Educational Technology Publication, 1974, Chapter 15.

Prais, S.J., and J. Aitchison "The Grouping of Observations in Regression Analysis." Revue of International Stat. Inst., 1954, 22: 1-22.

Raths, J. "The Appropriate Experimental Unit," Educational Leadership, 1967, 263-267.

Robinson, W.S. "Ecological Correlations and the Behavior of Individuals," American Sociological Review, 1950: 351-357.

Rock, D.A., L.L. Baird, and R.L. Linn "Interaction between College Effects and Student's Aptitudes," American Educational Research Journal, 1970, 7: 109-121.

Rogosa, D.R. "Measurement Error and the Johnson-Neyman Technique." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April, 1977.

Rogosa, D.R. "Aggregation, Quasi-experiments and Other Wonders." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Rogosa,D.R. "Causal Models in Longitudinal Research" To appear in Nesselroade J.R.

Rogosa, D.R. "Some Results for the Johnson-Neyman Technique," Unpublished Doctoral Dissertation, Stanford University, 1977.

Rosenshine, B.V. "Formal and Informal Teaching Styles; A Review of S.N. Bennett and Others, 'Teaching Styles and Pupil Progress,'" American Educational Research Journal, 1978 (in press).

Schmidt, W.C. "Covariance Structure Analysis of the Multivariate Random Effects Model." Unpublished Doctoral Dissertation, University of Chicago, 1969.

Shively, W.P. "Ecological Inference: The Use of Aggregate Data to Study Individuals." American Political Science Rev., 1969, 63: 1183-1196.

Stallings, J.A., and D.H. Kaskowitz "Follow Through Classroom Observation Evaluation 1972-1973, Stanford Research Institute, August, 1974.

Steck, J.C. "The Independence of Observations Obtained in Classroom Research," Unpublished

Stebbins, L.B. R.G. St. Pierre, E.C. Proper, R.B. Andersen, and T.R. Cerva Education as Experimentation: A Planned Variation Model, Vol IV-A An Evlauation of Follow Through, April, 1977.

Walberg, H.J and R. Singh "Teacher Quality Perceptions and Achievement in Rajasthan," The Alberta Journal of Educational REsearch, 1974, 20: 226-232.

Walsh, J.E. "Concerning the Effect of Intraclass Correlation on Certain Significance Tests." Annals of Math. Stat., 1947, 18: 88-96

Webb, N. "Learning in Individual and Small Group Settings." Unpublished Doctoral Dissertation, Stanford University, 1977.

Webb, N., F. Capell and F. Nielsen "Summary of Presentation and Discussion," Conference on Methodology for Aggregating Data in Educational Research, Stanford Universtiy, October, 1976, Technical Report No. 15, The Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez Associates, Ltd., October, 1976.

Weibull, M. "The Distribution of t- and f-statistics and of Correlations and Regression Coefficients in Stratified Samples from Normal Populations with Different Means." Skandinavisk Aktuarietidskrift, (Sup.) 1953, 36; 407-416.

Werts, C.E. and R.L. Linn "Considerations when Making Inferences withing the Analysis of Covariance Models." Education and Pyschological Measurement, 1971, 31: 407-416.

Werts, C.E. and R.L. Linn "A Regression Model for Compositional Effects," Unpublished paper. Princeton, N.J.: Educational Testing Service, 1969.

Wiley, D.E. and R.D. Bock "Quasi-experimentation in Educational Settings: Comment." The School Review, 1967, 75: 353-366.

Wittrock, M.D. and D.E. Wiley The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970.

Young, A.A. and L. Erbring "Unraveling Classroom Effects: A Multilevel Feedback Model of Contextual Effects." Technical Report No. 22, The Consortium on Methodology for Aggregating Data in Educational Research, Milwaukee, Wisc.: Vasquez Associates, Ltd., 1976.

Additional References

Burstein, L. "Some Methodological Considerations in the Analysis of Multilevel Data," Presentation at a Meeting of the May 12th Evaluation Group, Virginia, Polytechnic Institute, Blacksburg, VA. October, 1977.

Burstein, L. "Three Key Topics in Regression-Based Analysis of Multilevel Data from Quasi-Experiments and Field Studies". Paper presented at the Institute for Research on Teaching, Michigan State University, December, 1977.

Burstein, L., Fischer, K. and Miller, M.D. "Social Policy and the Effects of Social Background on Educational Performance: A Cross-National Comparison". Working Paper Series, Center for the Study of Evaluation, 1978 (draft).

Joreskog, K.G. "A General Method for Analysis of Covariance Structures". Biometrika, 1970, 57: 239-251.

McDonald, F.J. and P.J. Elias The Effects of Teaching Performance on Pupil Learning, BTES Phase II Final Report, Educational Testing Service, 1976.

Arora, S.S. "Error Components Regression Models and Their Applications." Annals of Economic and Social Measurement, 1973, 4:451-461.

Maddala, G.S. "The Use of Variance Components Models in Pooling Cross-Section and Time-Series of Crosssections," Econometrica, 1971, 39:359-382.

Blalock, H.M. Causal Inferences in Nonexperimental Research, New York: W.W. Norton and Company, 1964

Bloom, B.S. Toward a theory of testing which includes measurement-evaluation-assesment, in The Evaluation of Instruction: Issues and Problems, eds., M.C. Wittrock and D.E. Wiley, New York: Holt, Rinehart and Winston, Inc., 1970

Coleman, J.S. Methods and results in the IEA Studies of effects of school on learning, Review of Educational Research, 1975, 45:355-386.

Coleman, J.S., et al. Equality of Educational Opportunity. Washington DC.: U.S. Department of Health, Education and Welfare, U.S. Office of Education, OE-38001, 1966, pp. 218-333.

Comber, L.C. and Keeves, J.P. Science and Education in Nineteen Countries, International Studies in Evaluation, Vol. 1. Stockholm: Almqvist and Wiksell and New York: Wiley, 1973.

Peaker, G.F. An Empirical Study of Education in Twenty-One Countries. International Studies in Evaluation, Vol. VIII. Sweden: Almqvist & Wiksell, Publishers, 1974

Karweit, N. and Fennessey, J. A pragmatic framework for studying school effects: Estimation experiments using actual and simulated data, A progress report, Center for Social Organization of Schools, John Hopkins University, April, 1977.

Snow, R.E. Learning and individual differences, Review of Research in Education, 1976.

Wiley, D.E. Design and analysis of evaluation studies. In : Wittrock and Wiley (eds) The Evaluation Of Instruction: Issues and Problems. New York: Holt, Rinehart and Winston, 1970.

Light, R.J. and Smith, P.V. Accumulating evidence: Procedures for resolving contradictions among different research studies, Harvard Educational Review, 1971, 41:429-471.

Marco, G.L. A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 1974, 11: 225-234.

Glass, G.V. Primary, secondary, and metaanalysis of research, Educational Researcher, 1976, 5: 3-8.